Gene, 63 (1988) 261–275
Elsevier

GEN 02321

Nucleotide sequence of the *Drosophila* glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene

(Recombinant DNA; cDNA; library screening; dosage compensation; nucleotide sequencing; gene transformation; intron; exon; 'housekeeping' gene; phage λ vector)

David Fouts[a], Ranjan Ganguly[a]*, Anthony G. Gutierrez[b], John C. Lucchesi[b] and Jerry E. Manning[a]

[a] Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92717 (U.S.A.) Tel. (714)856-6034, and [b] Department of Zoology and Curriculum in Genetics, University of North Carolina, Chapel Hill, NC 27514 (U.S.A.) Tel. (919)962-1332

Received 20 August 1987
Revised 30 October 1987
Accepted 2 November 1987
Received by publisher 31 December 1987

SUMMARY

Glucose-6-phosphate dehydrogenase (G6PD) has a major role in NADPH production and is found in almost all cell types. The structural gene for G6PD is X-linked in *Drosophila melanogaster*, as it is in most eukaryotic organisms, and due to its ubiquitous expression, it can be considered a typical 'housekeeping' gene. Here we present the complete nucleotide (nt) sequence of *G6PD* cDNAs as well as the genomic copy of the *G6PD* gene. The *G6PD* gene has three introns so that the protein-coding region is divided into four segments. The 5′-end of mature *G6PD* mRNA is located 289 ± 1 nt upstream from the start codon. The sequence upstream from the transcription start point is G + T-rich and contains no commonly found transcription regulatory elements, such as a TATA box or GGGCGG sequence. *D. melanogaster* G6PD is 65% homologous with the human G6PD protein but has no homology with the human sequence for the first 42 amino acid residues. The G6PD gene was shown to be active when transduced to autosomal positions. For each transformant, G6PD activity in both male and female adults was not significantly different, indicating that the transduced gene, unlike the resident *G6PD*, is not dosage-compensated in males.

INTRODUCTION

Glucose-6-phosphate dehydrogenase (G6PD; D-glucose-6-phosphate : $NADP^+$ oxidoreductase, EC 1.1.1.49) is the first and dominant regulatory enzyme in the hexose monophosphate shunt. The primary role of G6PD is to generate NADPH, a reductant necessary for numerous physiological and

Correspondence to: Dr. J.E. Manning, Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92717 (U.S.A.) Tel. (714)856-5578.

* Present address: Department of Zoology and Cell, Molecular and Developmental Biology Program, University of Tennessee, Knoxville, TN 37996 (U.S.A.) Tel. (615)856-5578

Abbreviations: aa, amino acid(s); bp, base pair(s); DHFR, dihydrofolate reductase; G6PD, glucose-6-phosphate dehydrogenase; *G6PD*, gene coding for G6PD; kb, kilobases or 1000 bp; NADP, nicotinamide adenine dinucleotide phosphate; NADPH, reduced NADP; nt, nucleotide(s); *Zw*, Zwischenferment, *G6PD* gene in *Drosophila*.

biosynthetic processes (Beutler, 1983). This enzyme has been found in all organisms and cell types thus far analyzed, thus placing the *G6PD* gene in the category of general 'housekeeping' genes. G6PD activity has been used extensively for molecular, developmental, physiological and population studies with a variety of organisms, including *Drosophila* and man (Yoshida and Beutler, 1986). Recently, the gene for human and *Drosophila* G6PD have been cloned (Persico et al., 1986; Martini et al., 1986; Takizawa et al., 1986; Ganguly et al., 1985; Hori et al., 1985). The amino acid sequence of human G6PD has been determined by direct amino acid sequence analysis of the purified enzyme (Takizawa et al., 1986), and it has been derived from the sequence of cDNA clones which encode the enzyme (Persico et al., 1986; M.G. Persico, personal communication). Comparison of the two human amino acid sequences shows them to be identical except for a region at the N terminus of the protein, where no homology is present. No obvious explanation for the differences in sequence within this region of the protein is available.

Since common metabolic enzymes often show extensive amino acid sequence homology between humans and *Drosophila*, we wished to compare the amino acid sequence of *Drosophila* G6PD to that of human G6PD with the premise that such a comparison might provide insight into the discrepancy between the two human G6PD sequences. This report, therefore, presents the complete cDNA and genomic DNA sequences which encode G6PD in *Drosophila* and presents the surprising observation that the amino acid and DNA coding sequence of *Drosophila* G6PD diverges from that of human G6PD at the precise amino acid position that marks the start of sequence divergence between the two human G6PD sequences.

MATERIALS AND METHODS

(a) Isolation of cDNA clones

A cDNA library was constructed in phage λ gt10 using adult *Drosophila* poly(A)$^+$ RNA by methods described in Maniatis et al. (1982). The library was co-screened with two genomic DNA fragments containing the *G6PD* gene. One fragment contained sequences present in exon I (i.e., the 927 bp from *Pvu*II to *Bam*HI) and the other fragment included sequences present in the other three exons (i.e., the 1022 bp from *Eco*RI to *Pst*I). Approximately 600 000 independent recombinant phages were screened, and two recombinant phages, λDmC20 and λ-DmC21, showed positive hybridization with both probes. The inserts present in both phages were excised, subcloned into pUC9 and Bluescript, and a restriction enzyme map of both was constructed.

(b) Nucleotide sequence analysis

The strategy for determining the nucleotide sequence of the genomic and cDNA fragments which encode G6PD is shown in Fig. 1. The sequence was determined by the dideoxynucleotide chain-termination method of Sanger et al. (1977) and was verified by data generated from both strands. Growth and manipulation of phages M13mp18 and M13mp19 were as described previously (Messing, 1983). Sequential deletions of DNA fragments cloned in Bluescript with exonuclease III were performed using conditions suggested by the supplier (Stratagene, Inc.).

(c) S1 nuclease protection

A 408-bp *Sau*3A fragment (see Fig. 3), which spans exon I, was cloned into the *Bam*HI site of M13mp18. Recombinant phage DNA containing antisense strand was used as a template to synthesize ^{32}P-labeled sense-strand DNA. The *Sau*3A fragment was excised by digestion with *Eco*RI + *Sal*I and isolated by electroelution from a non-denaturing 5.5% polyacrylamide gel. The purified fragment was denatured and hybridized with 10 μg of adult *Drosophila* poly(A)$^+$ RNA for 16 h at 46°C in a 10-μl reaction mixture. As described previously (Casey and Davidson, 1977; Ganguly et al., 1985), these conditions allow RNA/DNA hybridization in the absence of DNA/DNA hybridization. Following hybridization, the RNA/DNA hybrids were treated with S1 nuclease and the size of the protected DNA fragment was resolved on a sequencing gel.

(d) Construction of *G6PD* Carnegie-20 transformation vector

A 6.7-kb *Hpa*I-*Sst*I DNA fragment (Ganguly et al., 1985) putatively containing the entire *G6PD* gene sequence was inserted into the single *Sal*I site of the transformation vector Carnegie-20 (Rubin and Spradling, 1982) following conversion of the *Hpa*I and *Sst*I sites into *Xho*I sites. First, the 4.9-kb *Hin*dIII-*Bam*HI fragment containing exons II–IV was cloned into pBR322. After *Sst*I cleavage, the *Sst*I site in the recombinant plasmid DNA was converted to an *Xho*I site by ligation with synthetic *Xho*I linkers. The resultant plasmid was designated pBrt. Second, the 5.8-kb *Eco*RI fragment containing exon I was cloned into pUC9, and the single *Hpa*I site was converted to an *Xho*I site using synthetic *Xho*I linkers. The two modified *G6PD*-containing DNA fragments were joined at the common *Eco*RI site in intron I. The 6.7-kb *Xho*I fragment (i.e., *Hpa*I to *Sst*I) was excised and inserted into the *Sal*I site of Carnegie-20. Orientation of the *G6PD* gene relative to the ry^+ gene was determined by digestion of the constructs with endonuclease *Hin*dIII.

(e) Transformation experiments

Germ-line transformations were performed following the procedures of Rubin and Spradling (1982; 1983) and Karess and Rubin (1984). The Carnegie-20 vector containing Zw^+ was microinjected (300 μg/ml) along with the helper plasmid pπ25.7wc (80 μg/ml) into preblastoderm embryos from a try^{506} or an Adh^{fn6} *cn* ry^{506} recipient strain. Injected embryos were reared to adulthood and mated to individuals from the same strain. Their progenies were examined for the presence of transformants detectable on the basis of their Ry^+ eye color phenotype. Separate lines were established by backcrossing individual transformants to flies of the appropriate sex from the recipient stock. To ensure that each transformant line represented a separate insertional event, only one transformant produced by a given recipient was retained. By monitoring the transmission of the Ry^+ phenotype in these lines, sex chromosome versus autosomal linkage of the transduced genes could be determined. A more precise cytological localization of these inserts was obtained by in situ hybridization. Plasmid pπ25. 7wc was labeled with a biotinylated deoxynucleotide (Bio-16 dUTP, Bethesda Research Laboratories, Gaithersburg, MD) and allowed to hybridize to its homologous sequences on larval salivary gland polytene chromosomes; its presence was detected by the binding of a streptavidin-biotin-horseradish peroxidase complex (ENZO Biochem. Inc., New York) according to a method modified by E. Hafen (personal communication). To measure the activity of transduced genes without the complication of an endogenous background, crosses were performed to replace the X chromosome of transformant lines with an X containing a Zw^- allele. We used Zw^{n1}, induced by ethylmethane sulfonate mutagenesis (Hughes and Lucchesi, 1977) and Zw^{H7a} recovered from a hybrid dysgenesis cross (Nero, 1987).

(f) Enzyme assays

Crude extracts were prepared by homogenizing adult males in 0.1 M Tris, 5 mM mercaptoethanol, 0.2 mM EDTA, 0.1 mM NADP buffer (pH 8.0) at a concentration of 10 mg of wet weight/ml. G6PD activity was measured as the increase in absorbance at 340 nm resulting from the reduction of NADP (Lucchesi and Rawls, 1973).

RESULTS

(a) Isolation of cDNA clones

Approximately 600000 λ recombinant phages from a cDNA library constructed from adult poly(A)$^+$ RNA were screened with DNA fragments containing exon I and exons II–IV of the genomic *G6PD* gene (Fig. 1). Two phages, designated λ-DmC20 and λDmC21, were identified and proved positive upon re-screening. The sizes of the cDNA inserts in λDmC20 and λDmC21 were assessed, by electrophoresis in a 1% agarose gel, to be 2000 bp and 1950 bp, respectively. Restriction enzyme analysis of these two cDNA inserts showed an almost totally overlapping pattern. The similarity of these two cDNAs was further confirmed by hybridization to a restriction enzyme digest of genomic *G6PD* sequences. Both cDNAs showed hybridization to DNA restriction fragments in exon I and exons

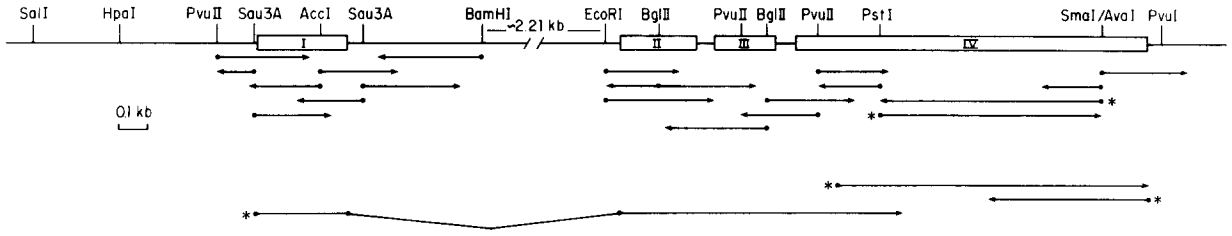


Fig. 1. Restriction map and sequencing strategy of the *G6PD* genomic and cDNA clones. Origins of arrows indicate the restriction sites used for sequencing. Arrows denoted by an asterisk were sequenced in both directions by sequential deletion with exonuclease III as described in MATERIALS AND METHODS, section **b**. The boxed regions denote the position of exons I–IV of the *G6PD* gene as determined by sequencing of *G6PD* cDNAs and S1 nuclease mapping of the transcriptional start point (see Fig. 3).

II–IV, but no hybridization was observed to any region of intron I.

(b) Sequence analysis

To more precisely define the number and position of exon/intron domains within the genomic *G6PD* gene, the complete nucleotide sequence of the cDNAs and the genomic DNA fragments containing exon I (i.e., *Pvu*II to *Bam*HI) and exons II–IV (i.e., *Eco*RI to *Pvu*I) were determined by the dideoxy chain-termination method (Fig. 2). The sequence of both complementary strands of all DNA fragments was determined and the position of 6-bp restriction enzyme sites predicted by either the sequence or restriction mapping was confirmed. The sequencing strategy is given in Fig. 1. The genomic and cDNA sequences are shown in Fig. 2 and underscored by the predicted amino acid sequence of the protein.

Direct comparison of the genomic and cDNA sequences confirms the position of intron I but also indicates the presence of two additional small introns in the previously designated exon II (Ganguly et al., 1985). Therefore, exon II must now be redefined as containing three exons and two introns. The three intron sequences which separate the four exons have several notable features. First, the junctions which define the intron/exon boundaries agree with the GT---AG rule of 5′-donor 3′-acceptor splice sites (Mount, 1982). Also present in the three introns are regions corresponding both in position and sequence to the consensus sequence (C/T)T(A/G)A(T/C) proposed as a 3′ splice signal in *Drosophila* (Keller and Noon, 1985). These authors also suggest that a second criterion for proper splicing is the absence of an AG between −3 and −19 from the 3′ splice point; a feature also shared by the three *G6PD* introns.

The sequence of the cDNA differed from the genomic DNA in three places. However, the change in nucleotides at these three sites did not alter the predicted amino acid sequence. This degree of difference is consistent with the extent of genetic polymorphism that might be anticipated, since the genomic DNA is from the Canton S strain while the cDNA originated from an Oregon R library.

The 3′ end of the gene was identified by the presence of the A residues at the 3′ terminus of the cDNAs for which no genomic counterpart was identified. Consistent with this assignment is the observation that both cDNAs showed terminal sequences which differed only in the number of the A residues between the *Eco*RI linker sequence and the putative polyadenylation site at nt position 2461. Also, the sequence ATTAAA at position 2426 resembles the consensus sequence AATAAA that precedes the polyadenylation site of most eukaryotic mRNAs by 12 to 30 nt residues. Since deviations from the consensus sequence have been described for several eukaryotic genes, in particular the chick actin gene (ATTAAA; Fornwald et al., 1986) and the human *G6PD* gene (ATTAAA; Persico et al., 1986), it is likely that the above sequence represents the polyadenylation signal for the *Drosophila G6PD* gene.

(c) 5′-End determination and transformation of *Zw*

To determine the 5′ end of *G6PD* mRNA, an S1 nuclease protection experiment was performed with the ^{32}P-labeled probe indicated in Fig. 3. The protected DNA fragment migrated as a single band of length 306 ± 1 nt on a nucleotide sequencing gel. This would position the 5′ end of the mRNA at nt −289 ± 1. In making this assignment we have

assumed that the terminal 92 nt of the probe that are present in intron I are not protected by mRNA. Also, since both strands of the probe are present during the hybridization reaction, it is important to note that under these hybridization conditions no protection of the probe (Fig. 3, lane 1) was observed in the absence of RNA.

Transformation experiments were conducted to determine the extent of the 5′ domain necessary to achieve gene activity. The P2OH2 Carnegie-20 vector contains the Zw^+-coding region flanked by 0.55 kb of upstream and 1.15 kb of downstream sequences. Crosses were performed to replace the X chromosome of transformant lines with an X containing a Zw^- allele so that the transduced gene could be measured without the complication of an endogenous background. All transduced genes were found to produce active G6PD enzyme although the level of enzyme activity differed substantially among transformant lines (ranging from 32% to 60%) due to position effects. To determine if *cis*-acting sequences responsible for dosage compensation were included in the Zw^+ sequences transduced to autosomal sites, males and females carrying a single dose of transduced genes were compared. Under these conditions, equal levels of activity in the two sexes signal the absence of compensation while higher levels of activity in males (ideally, twice as high as in females) indicate its occurrence. The results of these experiments, presented in Table II show the absence of compensation.

(d) Identification of the protein coding region

The first ATG in the G6PD transcript is 289 bp downstream from the 5′ end at position 1 in Fig. 2. Translation starting at this site would end at nt position 2277 (TGA) and would yield a 523-aa protein of M_r 60 100. A second ATG in the *G6PD* transcript is found at nt position 592 and is in the same reading frame as the first ATG. Initiation at the second ATG would result in a protein of 501 aa and an M_r of 57 676. Both of these predicted M_rs are in close agreement with the apparent M_r (i.e., 55 000) of the monomeric unit of G6PD (Lee et al., 1978; Williamson and Bentley, 1983). Examination of the 5′ sequences flanking the two potential translation start sites shows that neither strongly match the generalized *Drosophila* initiation consensus sequence proposed by Cavener (1987) (Table I). However, a comparison of these 5′ sequences with those sequences 5′ of the translational start sites of 83 other *Drosophila* genes reveals that the sequence 5′ of the first ATG is very similar to those of the glue protein Sgs-4 while those 5′ of the second ATG are similar to Hsp-22 (Table 1). Collectively, these observations provide no clear indication as to which ATG might serve as the predominant site for translational

TABLE I

Comparison of sequences 5′ of potential translation initiation sites for G6PD with consensus initiation sequences of *Drosophila* protein coding genes and the initiation sequence for the *Sgs-4* and *Hsp-22* genes

Consensus sequence[a]	a a a A a	t/c	C/A	A	A/C	A/C	ATG
First ATG[b]	T C G G G	T	C	A	A	G	ATG
Sgs-4[c]	C A A A G	T	C	A	A	G	ATG
Second ATG[d]	G T C G C	C	T	A	C	A	ATG
Hsp-22[e]	A T C A A	C	T	A	C	A	ATG

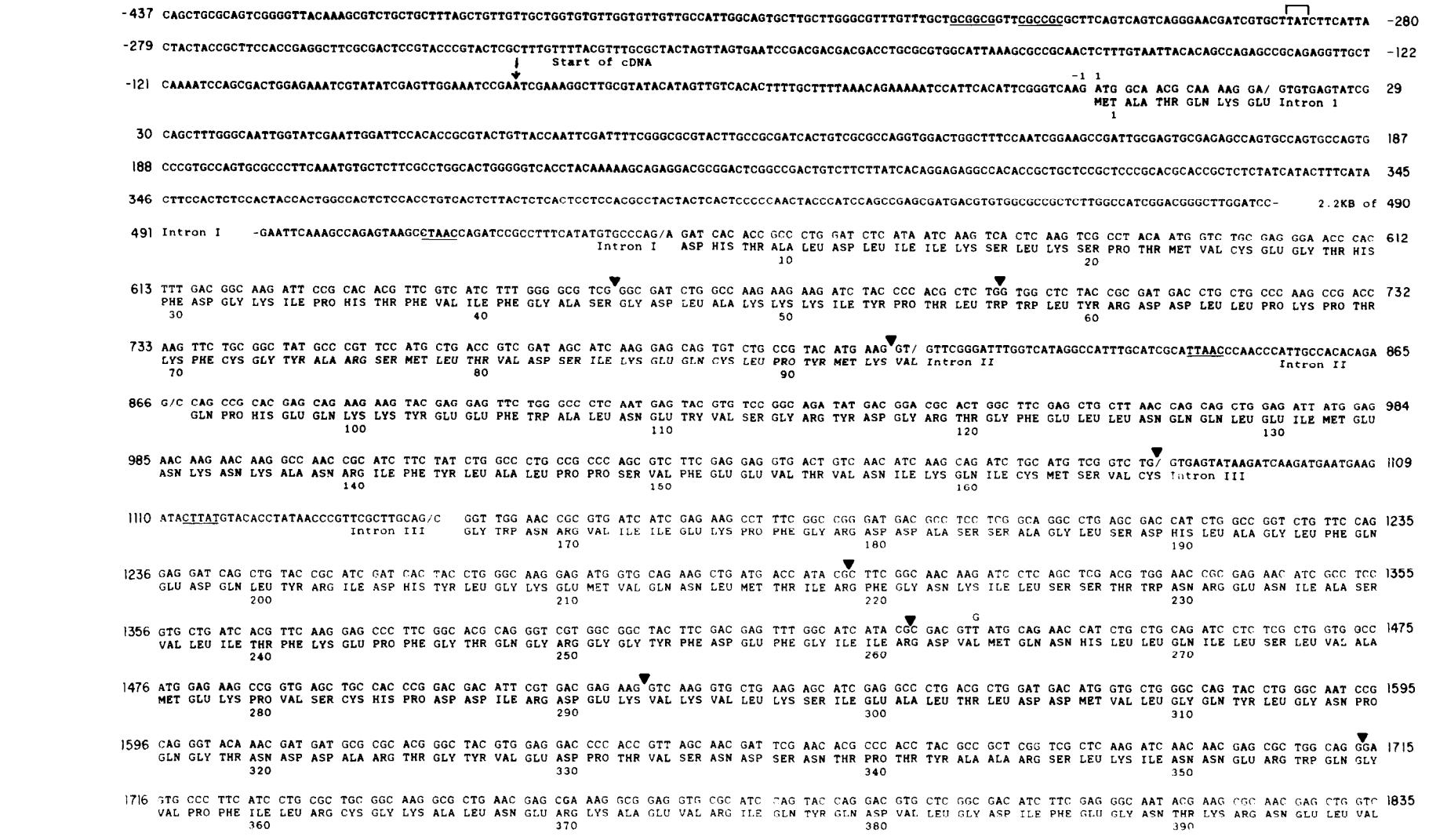
[a] The rules used for assignment of consensus are as follows (Cavener, 1987). If, in a compilation of *Drosophila* sequences flanking the translational start site, the frequency of a single nucleotide at a specific position is greater than 50% and greater than twice the number of the second most frequent nucleotide it is considered as the consensus nucleotide (upper-case letter). If the sum of the frequencies of two nucleotides is greater than 75% (but neither meets the criteria for a single nucleotide assignement) they are considered as co-consensus nucleotides. If no single nucleotide or pair of nucleotides meets the criteria of consensus nucleotide(s) the most frequent nucleotide is considered as the preferred nucleotide and is denoted as such by a lower-case letter.

[b] Nucleotides −10 to 3 in Fig. 2. Nucleotides −10 to −1 are 5′ to the first ATG in the *G6PD* transcript.

[c] The ATG initiation codon and the 10 nt immediately 5′ to this initiation site in the *Drosophila* glue protein gene, *Sgs-4* (Muskavitch and Hogness, 1982).

[d] Nt 582 to 594 in Fig. 2. Nt 582 to 591 are 5′ to the second ATG in the *G6PD* transcript.

[e] The ATG initiation codon and the 10 nt immediately 5′ to this initiation site in the *Drosophila* heat-shock protein gene, *Hsp-22* (Ingolia and Craig, 1981).


-437 CAGCTGCGCAGTCGGGGTTACAAAGCGTCTGCTGCTTTAGCTGTTGTTGCTGGTGTGTTGGTGTTGTTGCCATTGGCAGTGCTTGCTTGGGCGTTTGTTTGCTGCGGCGGTTCGGCCGCGCTTCAGTCAGTCAGGGAACGATCGTGCTTATCTTCATTA -280
-279 CTACTACCGCTTCCACCGAGGCTTCGCGACTCCGTACCCGTACTCGCTTTGTTTTACGTTTGCGCTACTAGTTAGTGAATCCGACGACGACGACCTGCGCGTGGCATTAAAGCGCCGCAACTCTTTGTAATTACACAGCCAGAGCCGCAGAGGTTGCT -122
Start of cDNA
-121 CAAAATCCAGCGACTGGAGAAATCGTATATCGAGTTGGAAATCCGAATCGAAAGGCTTGCGTATACATAGTTGTCACACTTTTGCTTTTAAACAGAAAAATCCATTCACATTCGGGTCAAG ATG GCA ACG CAA AAG GA/ GTGTGAGTATCG 29
-1 1
MET ALA THR GLN LYS GLU Intron 1
1
30 CAGCTTTGGGCAATTGGTATCGAATTGGATTCCACACCGCGTACTGTTACCAATTCGATTTTCGGGCGCGTACTTGCCGCGATCACTGTCGCGCCAGGTGGACTGGCTTTCCAATCGGAAGCCGATTGCGAGTGCGAGAGCCAGTGCCAGTGCCAGTG 187
188 CCCGTGCCAGTGCGCCCTTCAAATGTGCTCTTCGCCTGGCACTGGGGGTCACCTACAAAAAGCAGAGGACGCGGACTCGGCCGACTGTCTTCTTATCACAGGAGAGGCCACACCGCTGCTCCGCTCCCGCACGCACCGCTCTCTATCATACTTTCATA 345
346 CTTCCACTCTCCACTACCACTGGCCACTCTCCACCTGTCACTCTTACTCTCACTCCTCCACGCCTACTACTCACTCCCCCAACTACCCATCCAGCCGAGCGATGACGTGTGGCGCCGCTCTTGGCCATCGGACGGGCTTGGATCC- 2.2KB of 490
491 Intron I -GAATTCAAAGCCAGAGTAAGCCTAACCAGATCCGCCTTTCATATGTGCCCAG/A GAT CAC ACC GCC CTG GAT CTC ATA ATC AAG TCA CTC AAG TCG CCT ACA ATG GTC TGC GAG GGA ACC CAC 612
Intron I ASP HIS THR ALA LEU ASP LEU ILE ILE LYS SER LEU LYS SER PRO THR MET VAL CYS GLU GLY THR HIS
10 20
613 TTT GAC GGC AAG ATT CCG CAC ACG TTC GTC ATC TTT GGG GCG TCG GGC GAT CTG GCC AAG AAG AAG ATC TAC CCC ACG CTC TGG TGG CTC TAC CGC GAT GAC CTG CTG CCC AAG CCG ACC 732
PHE ASP GLY LYS ILE PRO HIS THR PHE VAL ILE PHE GLY ALA SER GLY ASP LEU ALA LYS LYS LYS ILE TYR PRO THR LEU TRP TRP LEU TYR ARG ASP ASP LEU LEU PRO LYS PRO THR
30 40 50 60
733 AAG TTC TGC GGC TAT GCC CGT TCC ATG CTG ACC GTC GAT AGC ATC AAG GAG CAG TGT CTG CCG TAC ATG AAG GT/ GTTCGGGATTTGGTCATAGGCCATTTGCATCGCATTAACCCAACCCATTGCCACACAGA 865
LYS PHE CYS GLY TYR ALA ARG SER MET LEU THR VAL ASP SER ILE LYS GLU GLN CYS LEU PRO TYR MET LYS VAL Intron II Intron II
70 80 90
866 G/C CAG CCG CAC GAG CAG AAG AAG TAC GAG GAG TTC TGG GCC CTC AAT GAG TAC GTG TCC GGC AGA TAT GAC GGA CGC ACT GGC TTC GAG CTG CTT AAC CAG CAG CTG GAG ATT ATG GAG 984
GLN PRO HIS GLU GLN LYS LYS TYR GLU GLU PHE TRP ALA LEU ASN GLU TRY VAL SER GLY ARG TYR ASP GLY ARG THR GLY PHE GLU LEU LEU ASN GLN GLN LEU GLU ILE MET GLU
100 110 120 130
985 AAC AAG AAC AAG GCC AAC CGC ATC TTC TAT CTG GCC CTG CCG CCC AGC GTC TTC GAG GAG GTG ACT GTC AAC ATC AAG CAG ATC TGC ATG TCG GTC TG/ GTGAGTATAAGATCAAGATGAATGAAG 1109
ASN LYS ASN LYS ALA ASN ARG ILE PHE TYR LEU ALA LEU PRO PRO SER VAL PHE GLU GLU VAL THR VAL ASN ILE LYS GLN ILE CYS MET SER VAL CYS Intron III
140 150 160
1110 ATACTTATGTACACCTATAACCCGTTCGCTTGCAG/C GGT TGG AAC CGC GTG ATC ATC GAG AAG CCT TTC GGC CGG GAT GAC GCC TCC TCG GCA GGC CTG AGC GAC CAT CTG GCC GGT CTG TTC CAG 1235
Intron III GLY TRP ASN ARG VAL ILE ILE GLU LYS PRO PHE GLY ARG ASP ASP ALA SER SER ALA GLY LEU SER ASP HIS LEU ALA GLY LEU PHE GLN
170 180 190
1236 GAG GAT CAG CTG TAC CGC ATC GAT CAC TAC CTG GGC AAG GAG ATG GTG CAG AAG CTG ATG ACC ATA CGC TTC GGC AAC AAG ATC CTC AGC TCG ACG TGG AAC CGC GAG AAC ATC GCC TCC 1355
GLU ASP GLN LEU TYR ARG ILE ASP HIS TYR LEU GLY LYS GLU MET VAL GLN ASN LEU MET THR ILE ARG PHE GLY ASN LYS ILE LEU SER SER THR TRP ASN ARG GLU ASN ILE ALA SER
200 210 220 230
G
1356 GTG CTG ATC ACG TTC AAG GAG CCC TTC GGC ACG CAG GGT CGT GGC GGC TAC TTC GAC GAG TTT GGC ATC ATA CGC GAC GTT ATG CAG AAC CAT CTG CTG CAG ATC CTC TCG CTG GTG GCC 1475
VAL LEU ILE THR PHE LYS GLU PRO PHE GLY THR GLN GLY ARG GLY GLY TYR PHE ASP GLU PHE GLY ILE ILE ARG ASP VAL MET GLN ASN HIS LEU LEU GLN ILE LEU SER LEU VAL ALA
240 250 260 270
1476 ATG GAG AAG CCG GTG AGC TGC CAC CCG GAC GAC ATT CGT GAC GAG AAG GTC AAG GTG CTG AAG AGC ATC GAG GCC CTG ACG CTG GAT GAC ATG GTG CTG GGC CAG TAC CTG GGC AAT CCG 1595
MET GLU LYS PRO VAL SER CYS HIS PRO ASP ASP ILE ARG ASP GLU LYS VAL LYS VAL LEU LYS SER ILE GLU ALA LEU THR LEU ASP ASP MET VAL LEU GLY GLN TYR LEU GLY ASN PRO
280 290 300 310
1596 CAG GGT ACA AAC GAT GAT GCG CGC ACG GGC TAC GTG GAG GAC CCC ACC GTT AGC AAC GAT TCG AAC ACG CCC ACC TAC GCC GCT CGG TCG CTC AAG ATC AAC AAC GAG CGC TGG CAG GGA 1715
GLN GLY THR ASN ASP ASP ALA ARG THR GLY TYR VAL GLU ASP PRO THR VAL SER ASN ASP SER ASN THR PRO THR TYR ALA ALA ARG SER LEU LYS ILE ASN ASN GLU ARG TRP GLN GLY
320 330 340 350
1716 GTG CCC TTC ATC CTG CGC TGC GGC AAG GCG CTG AAC GAG CGA AAG GCG GAG GTG CGC ATC CAG TAC CAG GAC GTG CTC GGC GAC ATC TTC GAG GGC AAT ACG AAG CGC AAC GAG CTG GTC 1835
VAL PRO PHE ILE LEU ARG CYS GLY LYS ALA LEU ASN GLU ARG LYS ALA GLU VAL ARG ILE GLN TYR GLN ASP VAL LEU GLY ASP ILE PHE GLU GLY ASN THR LYS ARG ASN GLU LEU VAL
360 370 380 390

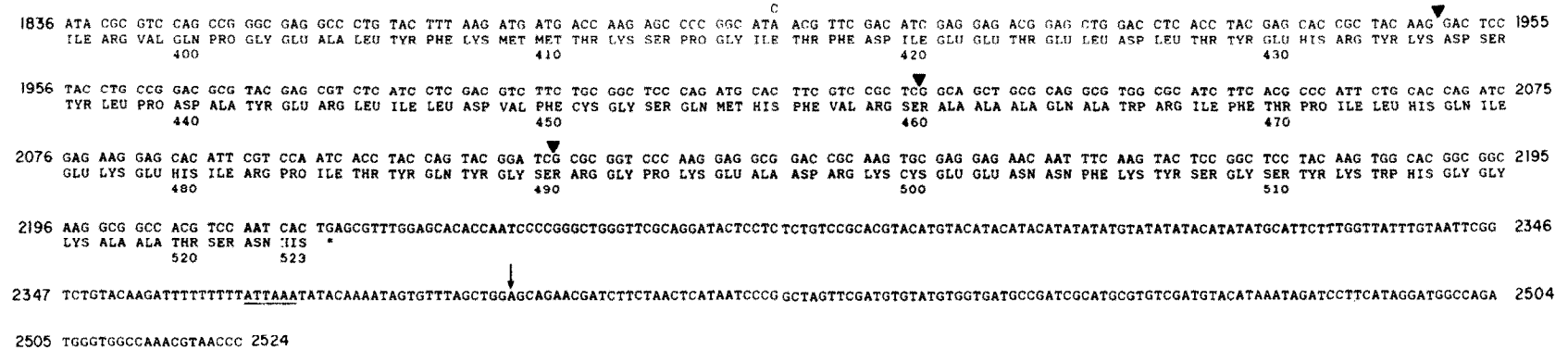


Fig. 2. The nucleotide sequence of the *Drosophila G6PD* gene. The nucleotide sequence encompassing the *G6PD* gene starts at the *Pvu*II site and extends 66 nt downstream from the *Pvu*I site shown in Fig. 1. The proposed amino acid sequence of G6PD is given below the nucleotide sequence. Nucleotides 5′ upstream from the proposed translational initiation codon are given negative numbers, starting with −1. Nucleotides 3′ downstream from the start codon are given positive numbers starting at the A base in the proposed ATG translational initiation codon. Numbering of amino acids starts at the Met within exon I. Underlining marks G + C-rich inverted repeat (in the first line) 5′ upstream from the translation start site and the polyadenylation signal at the 3′ end (in the second to the last line). Underlining within intron regions denotes putative splice signals. The horizontal bracket marks the putative transcriptional start point(s) as determined by S1 nuclease analysis (see RESULTS, section **c**, and Fig. 3), and the downward thin arrow marks the poly(A) attachment site. The solid triangles denote the position of intron sequences in the human *G6PD* gene. Single nucleotides placed above the gene sequence (Canton S) are nucleotide changes found in the cDNA clones of Oregon R. The asterisk marks the position of the putative stop codon in exon IV. The broken arrow in line two marks the 5′-most nucleotide in the cDNA present in λ Dmc20. Nucleotide sequences found 5′ of this point were determined by sequencing genomic DNA fragments.

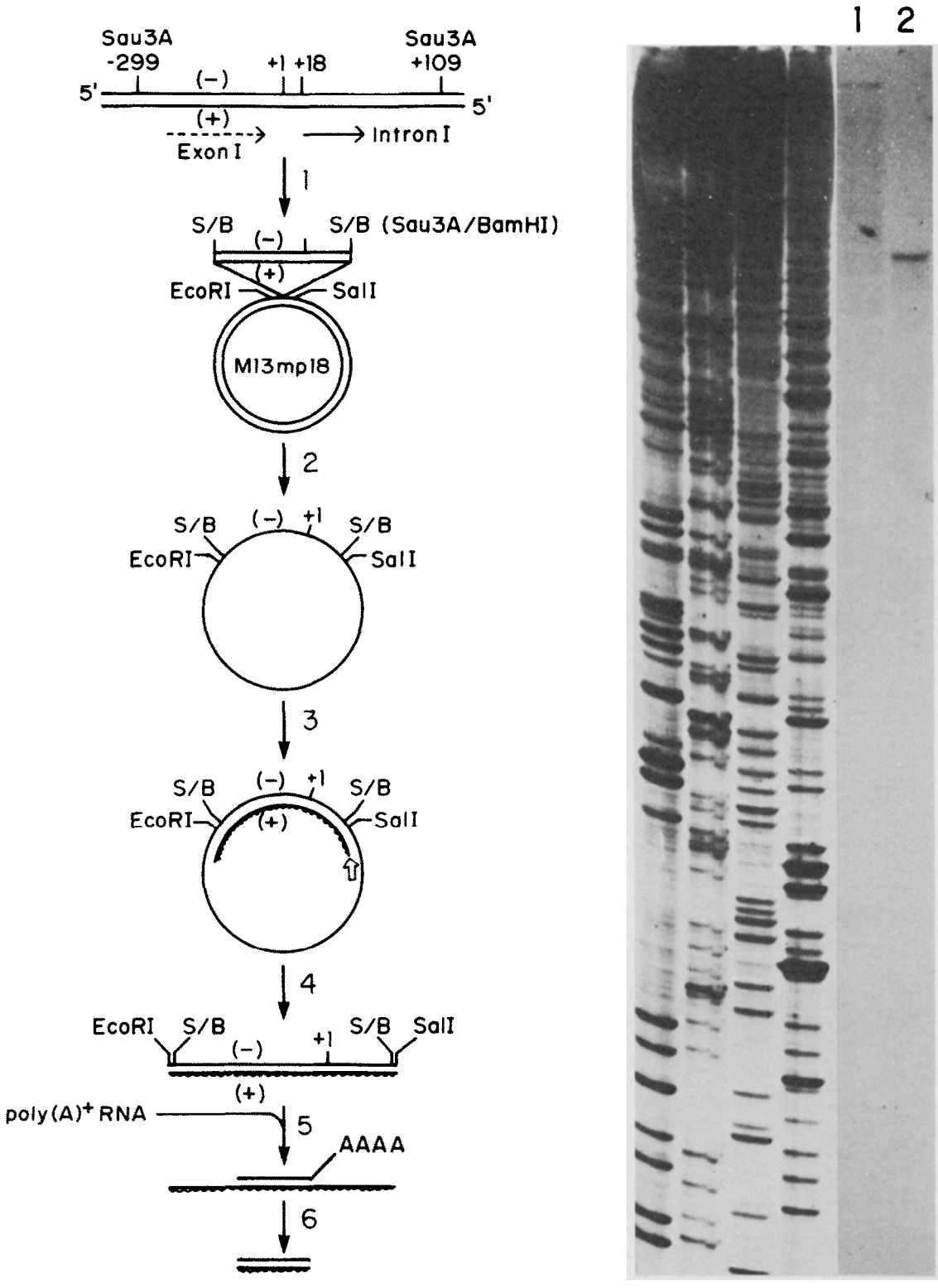


Fig. 3. S1 nuclease analysis of the 5′ end of *G6PD* mRNA. A schematic diagram showing the approach used for mapping the 5′ end of the G6PD mRNA is shown in the left column of the figure. Step 1 is the insertion of a 408-bp *Sau*3A fragment, which spans exon *I*, into the *Bam*HI site of M13mp18. Single-stranded recombinant phage DNA containing antisense strand (−) was isolated (step 2) and used as a template to synthesize ^{32}P-labeled sense-strand DNA (step 3). The *Sau*3A fragment containing the ^{32}P label in the sense

TABLE II

Comparison of G6PD transformed and control males and females carrying a single dose of an active Zw^+ gene

Strain	Chromosomal site[d]	Ratio of G6PD activity (males/females)[e]		
		Mean	S.D.	*n*
Control 1[a]	18D	2.03	0.26	(4)
Control 2[b]	18D	1.89	0.23	(4)
H2-248M[c]	69A	1.04	0.26	(3)
H2-263M[c]	84EF	0.85	0.25	(3)
H2-218F[c]	47C	0.99	—	(2)

[a] $+/Y$; ry^{506}/ry^{506} and $+/y\ sc\ cv\ v\ Zw^{H7a}$; ry^{506}/ry^{506}

[b] $+/Y$; $ry^{506}/+$ and $+/y\ sc\ cv\ v\ Zw^{H7a}$; $ry^{506}/+$

[c] $cv\ sc\ cv\ v\ Zw^{H7a}/Y$; $ry^{506}/ry^{506}[ry + Zw^+]$ and $y\ sc\ cv\ v\ Zw^{H7a}/y\ sc\ cv\ v\ Zw^{H7a}$; ry^{506}/ry^{506} $[ry^+ Zw^+]$.

[d] The chromosomal sites of the indigenous or transduced Zw^+ genes are given using the notations of the larval salivary gland chromosome map of LeFevre (1976).

[e] G6PD levels were expressed in units of activity, where one unit is the activity necessary to reduce 1.0 μmol of NADP/mg of live weight/min. The ratio of the activity in an extract from males to that in an extract from females was calculated. *n* is the number of independent determinations of this ratio for a given control or experimental line. Presented in the table are the means of the ratios and their standard deviation (S.D.).

initiation. Therefore, we have tentatively assigned the first ATG triplet as the start codon, because the most upstream ATG in a reading frame is used most frequently for initiation (Kozak, 1984).

(e) Nucleotide homology between *Drosophila* and human G6PD-coding sequences

A homology matrix analysis of nucleotide identities between the *Drosophila* and human cDNA sequences (Persico et al., 1986; M.G. Persico, personal communication) is shown in Fig. 4. Sequences in the 5′ end upstream of *Drosophila* nt 649 and in the 3′ untranslated region of the two genes are not shown because no substantial regions of homology are present. Position homology between the two sequences is scored by placement of a symbol if 22 out of 31 nt showed identities in a sequential scan of the gene sequence. Most of the scored homologies lie on a single continuous linear axis with only a few regions being identified elsewhere within the sequence. The high degree of homology observed in this alignment (i.e., about 60%) indicates that the coding region of the two genes has been conserved in length with few, if any, insertions, deletions or gene rearrangements.

Direct comparison of the two coding sequences shows sequence homology on both sides of *Drosophila* intron II and intron III. *Drosophila* intron III is in the same position as human intron V and *Drosophila* intron II starts two bases 3′ of the position of human intron IV. No sequence homology is observed between the *Drosophila* intron sequences and the corresponding partial sequences in the human gene. Human intron II and introns VI–XII have no counterpart in the *Drosophila* gene. However, the positions of these intron sequences in the human gene all occur within regions of almost precise homology with the *Drosophila* sequence. Human intron I is in the 5′-nontranslated leader sequence and has no counterpart in the *Drosophila* gene.

(f) Amino acid homology between the *Drosophila* and human G6PD protein

A comparison of the predicted amino acid sequence of the *Drosophila* G6PD protein and the human protein is shown in Fig. 5. Residues 1–53 in the human protein and aa 1–41 in the *Drosophila* protein (Takizawa et al., 1986) are not shown because no homology is observed in the amino acid sequences 5′ upstream from the Gly residue at aa

strand was excised by digestion with *Eco*RI + *Sau*3A and isolated by electroelution from a non-denaturing 5.5% polyacrylamide gel (step 4). The isolated fragment was denatured and hybridized with 10 μg of adult *Drosophila* poly(A)$^+$ RNA for 16 h at 46°C in a 10-μl reaction mixture (step 5). In a control experiment hybridization was performed in the absence of RNA. The hybrids were treated with S1 nuclease and their size(s) were determined by electrophoresis on a 6% sequencing gel (right side of figure). The gel was autoradiographed for either ten days (lanes 1 and 2) or two days (sequencing reaction in four lanes to the left). No signal was observed in the reaction without RNA (lane 1) while a single band of a length of 306 ± 1 nt was observed in the reaction containing poly(A)$^+$ RNA (lane 2). Assuming the terminal 92 nt on the 3′ end of the *Sau*3A fragment (i.e., intron I) are not protected by RNA, this result places the 5′ end of the RNA 10 bp downstream from the 5′ end of the *Sau*3A fragment at nt −289 ± 1 of the G6PD sequence (Fig. 2). Single-stranded DNA is represented as a single solid line. Double-stranded DNA is represented by a double solid line, and ^{32}P-labeled DNA is represented by a serrated line.

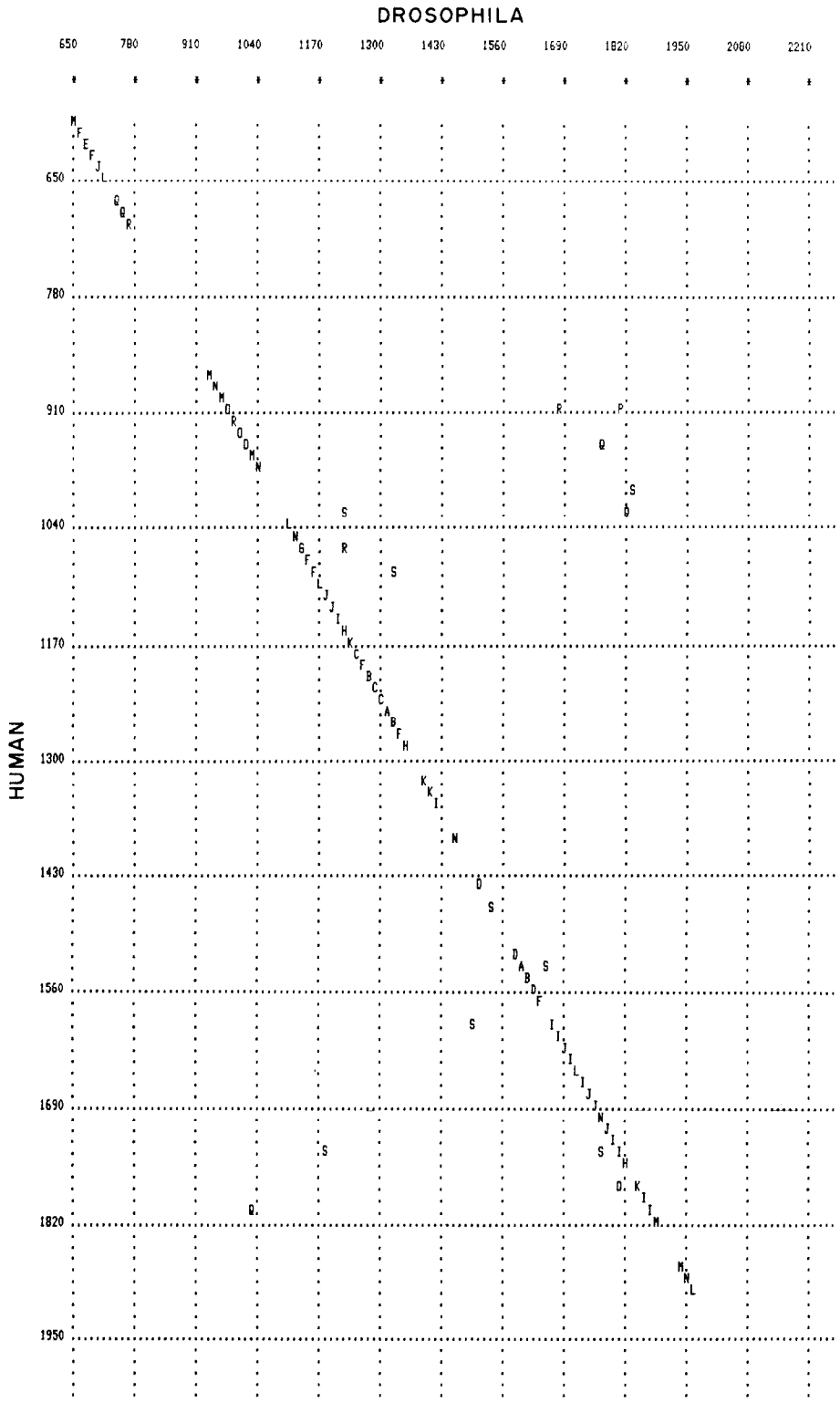


Fig. 4. Dot matrix comparison of *Drosophila* and human *G6PD* nucleotide sequences. Edited cDNA sequences encoding *Drosophila* and human *G6PD* are compared. Windows of 31 nt are sequentially compared and a letter scored for any 21 identical nt (Pustell and Kafatos, 1982; 1984; DNA/Protein Sequence Analysis System, International Biotechnologies, Inc., New Haven, CT). The letters represent varying degrees of homology among the 31 nt being scored. The letter A represents a match value of 100%, the letter B a value of 99–98%, with a progressive decrease in % homology to the letter S which represents 64–65% homology, or 21 of 31 nt having an identical match. The *Drosophila* sequence starts at the Gly residue at nt 649 and ends at the TGA stop codon (nt 2017). The human sequence starts at the ATG start codon and ends at the stop codon TGA (Persico et al., 1986).

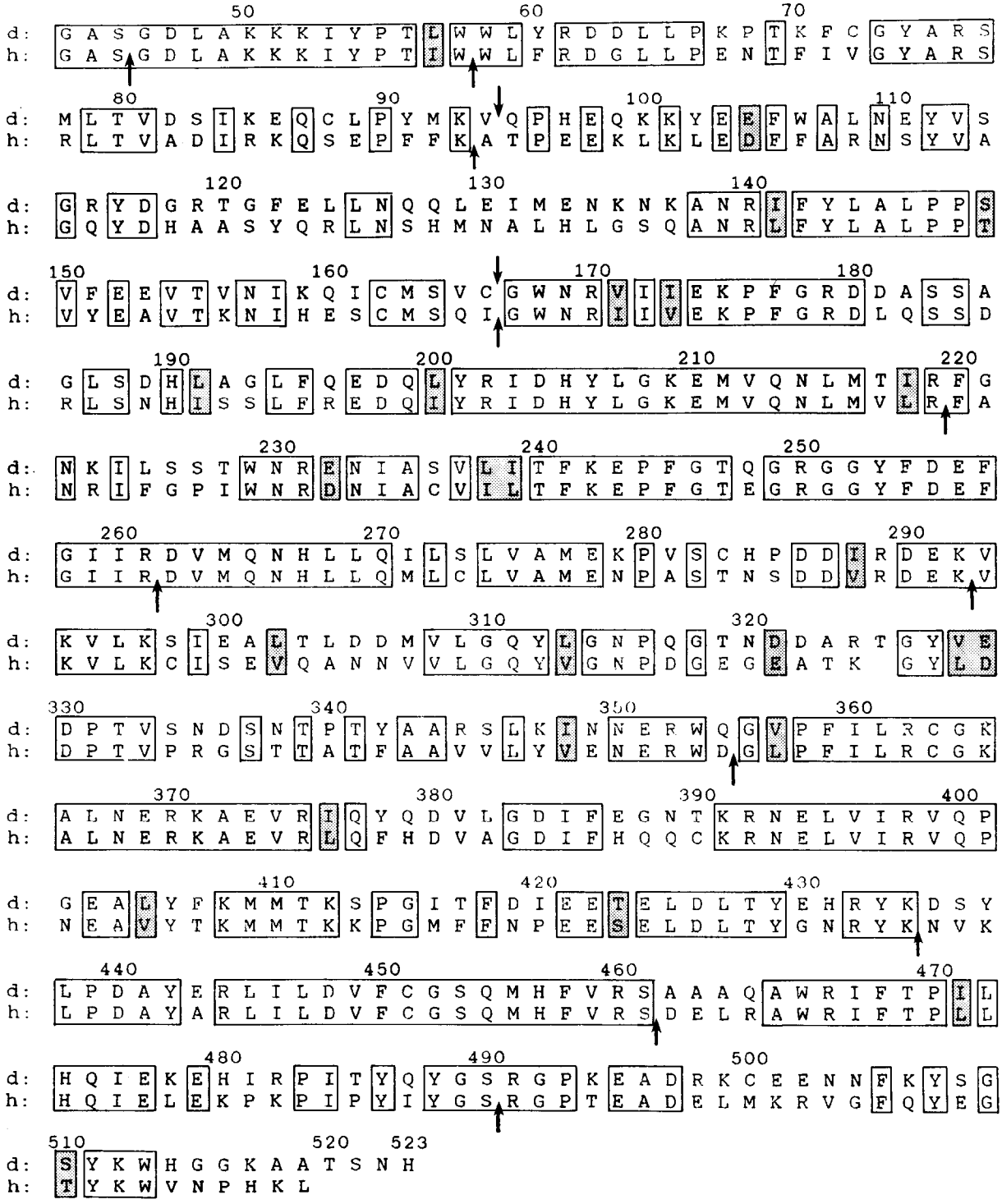


Fig. 5. Comparison of the *Drosophila* (d:) and the human (h:) G6PD amino acid sequences. In the comparison of the *Drosophila* and human G6PD sequence the first 41 aa at the N terminus of the *Drosophila* protein and the first 53 aa of the human protein (Takizawa et al., 1986) show no homology; therefore, for the clarity of presentation they were not listed. Adjustment of the sequences for alignment showing maximal homology required a one-codon shift of the human sequence at the aa residue 326 and an excision of a Glu residue in the human sequence between aa residues 465 and 466. Identical amino acid residues between the two sequences are denoted by open boxes. Conserved substitution of residues are denoted by shaded boxes, where I = L = V, D = E, S = T and F = Y. Downward arrows mark the position of introns in the *Drosophila* sequence, and upward arrows mark the position of introns in the human sequence. Numbering of the amino acid residues is identical to that shown in Fig. 2 and starts at aa position 42 in the *Drosophila* G6PD protein.

positions 42 and 54 of the *Drosophila* and human proteins, respectively. Also, for this N terminus region of the human protein the amino acid sequence predicted by the cDNA (Persico et al., 1986; M.G. Persico, personal communication) and the amino acid sequence obtained by direct protein sequencing (Takizawa et al., 1986) are dissimilar. Maximal alignment of the two protein sequences was obtained by displacement of one codon in the human protein at aa position 326 and excision of one codon in the human sequence at aa position 465. Approximately 63% of the amino acids are conserved using this alignment; if substitutions of amino acids with similar chemical properties are considered, the homology increases to 68%.

Considerable homology exists throughout the sequence, however, three regions of particularly strong homology (i.e., greater than 79%) are apparent. The first is near the N-terminal region (*Drosophila* aa residues 42–81); the second and third are near the central portion of the protein (*Drosophila* aa residues 193–318 and 351–460, respectively). The homology sharply decreases near the C terminus where the *Drosophila* sequence contains an additional 4 aa.

DISCUSSION

(a) Comparison of data

Hori et al. (1985) have reported the cloning of the *Drosophila* Zw^+ gene utilizing oligodeoxynucleotide probes derived from the amino acid sequence of a hexapeptide of *Drosophila melanogaster* G6PD. Surprisingly, when the sequence of the hexapeptide is compared with the complete amino acid sequence shown in Fig. 2, its position cannot be found. Neither can we identify, in the total nucleotide sequence of the *G6PD* gene, the sequence of the synthetic nucleotide probes used to isolate the *G6PD* gene. We find this particularly puzzling since the restriction map reported by Hori et al. (1985) is identical to that previously published by Ganguly et al. (1985). Although we have no explanation for this discrepancy, it seems likely that the data presented in Fig. 2 represent the complete nucleotide sequence of *Drosophila G6PD*. We base this assertion on the following arguments. The sequence shown in Fig. 2 is derived both from genomic DNA fragments which have previously been shown to encode G6PD (Ganguly et al., 1985) and from cDNA sequences which are homologous to the genomic DNAs. Furthermore, both the nucleotide sequence and the amino acid sequence inferred from the nucleotide sequence show extensive homology with human G6PD (Figs. 4 and 5; Persico et al., 1986; Takazawa et al., 1986). Finally, the genomic DNA fragments selected for nucleotide sequence analysis clearly contain the entire *G6PD* gene as evidenced by the results of the transformation experiments.

(b) Comparison of the human and *Drosophila G6PD* genes

One of the most interesting results from the comparison of the two G6PD sequences is the change in homology starting at the Gly residue at *Drosophila* nt 649. 3′ downstream from the Gly residue the two genes and their respective amino acid sequences show extensive homology. The similarities between the two genes within this region also extend to introns being found in common positions as shown in Fig. 2. These observations are, therefore, consistent with the thought that these two genes share a common ancestor. What is intriguing, then, is the concurrent loss of amino acid and nucleotide sequence homology between the *Drosophila* and the human G6PDs at the Gly residue and between the two human G6PDs at a Met residue which is immediately 5′ of the Gly residue. Within this 5′ region no homology is seen between the *Drosophila* and human *G6PD* sequence (Persico et al., 1986), and no similarities are observed between the amino acid sequence of *Drosophila* G6PD and that reported for human G6PD (Takazawa et al., 1986). Furthermore, the amino acid sequence of human G6PD 5′ of the Met residue as inferred by the cDNA sequence (M.G. Persico, personal communication) is non-homologous to the human G6PD sequence obtained by direct protein sequence analysis (Takazawa et al., 1986).

It is possible that the divergence of the three G6PD sequences at precisely the same amino acid is merely fortuitous and reflects some technical difficulty in either the nucleotide or amino acid sequence data. This seems unlikely, however, since the two

human sequences are identical in their central and 3′ regions and extensive homology with the *Drosophila* sequence is also found in these regions. A more plausible explanation is that the sequence differences in the human G6PD involve alternate splicing events at the 5′ termini of the messenger RNA. Recent studies on the messenger RNAs which encode human tyrosine hydroxylase reveal that, in man, tyrosine hydroxylase is encoded by three distinct messenger RNAs. Like the human G6PD sequences, these mRNAs and the proteins they encode are identical in their central and 3′ regions but diverge at their 5′ ends (Grima et al., 1987). Apparently, the heterogeneity at the 5′ end results from alternative splicing events within the primary transcript. If, in fact, multiple 5′ termini of human G6PD are the result of alternate splicing events, the extensive homology between the human and *Drosophila* genes in their central and 3′ regions suggests that the 5′ *Drosophila* sequences may yet be found within the human *G6PD* gene.

(b) Analysis of the promoter region

The S1 mapping experiment places the transcription start point 289 ± 1 nt upstream from the first possible translation start site. The region upstream from the transcription start point is very G + T-rich and shows some sequence moieties similar to PolII promoter regions. The sequence C–C–A–T–T, which differs by 1 nt from the canonical promoter sequence C–C–A–A–T, is found 75 bp upstream from the transcription start point. The *G6PD* gene lacks, however, the 'TATA' box which seems to position the RNA polymerase for accurate initiation and is normally found 20 to 30 nt upstream from the transcription start point. Although most genes containing PolII promoters possess a 'TATA' box, there are PolII promoters that lack the TATA box (Nevins, 1983). In particular, many housekeeping genes, i.e., genes which are fairly uniformly expressed in most tissue types throughout the life cycle of the organism, do not possess a TATA box: the hydroxymethyl glutaryl CoA reductase gene (Reynolds et al., 1984), the hypoxanthine phosphoribosyltransferase gene (Patel et al., 1986; Melton et al., 1984), the adenosine deaminase gene (Valerio, 1985), the DHFR gene (Masters and Attardi, 1985; Mitchell et al., 1986), one of the two glyceraldehyde-3-phosphate dehydrogenase genes in *Drosophila* (Tso et al., 1985), as well as the PrP 27-30 gene (Basler et al., 1986) and the U1 RNA gene (Roebuck and Stump, 1985). These genes do, however, contain one or more copies of the sequence GGGCGG or its inverse complement CCGCCC upstream from their transcription start point. This sequence is found four times in the mouse DHFR promoter (Dynan et al., 1986) and has been shown to be an important component of the SV40 virus early promoter (Barrera-Sladana et al., 1985), as well as the thymidine kinase promoter of *Herpes simplex* virus (McKnight et al., 1984). Examination of the sequence in Fig. 2 reveals that the above sequence does not appear in the first 148 nt upstream from the transcription start point of the *G6PD* gene. However, the sequence GCGGCG and its inverse complement CGCCGC are found 39 and 30 nt, respectively, upstream from the transcription start point. Although no function can be described to these sequences, their location and similarity to the G + C-rich promoter sequence described above invites the possibility that these sequences may be important in potentiation of transcription of the *G6PD* gene.

(c) Dosage compensation

The absence of dosage compensation of the Zw^+ gene relocated to an autosomal site is of some interest. Dosage compensation, i.e., the equalization of X-linked gene products in males and females, is achieved in *Drosophila* by an enhancement of transcription of X-linked genes in males. The *cis*-acting sequences responsible for this effect can be very closely linked to the coding portion of the gene, as in the case of w^+ (Levis et al., 1985; Pirrotta et al., 1985). They may even occur within the gene: a sequence located in the first intron of the per^+ gene allows it to remain compensated when it is relocated to an autosomal site (J. Hall, personal communication). In contrast to these cases, the *cis*-acting sequences responsible for the compensation of Zw^+ must be located further away from the coding portion of the gene than 0.55 kb of upstream and 1.15 kb of downstream sequences.

ACKNOWLEDGEMENTS

We thank D. Nero for providing us with the dysgenic Zw^- mutant and J. Massey for technical assistance. This investigation was supported by Grant GM15691 awarded by the National Institute of Health.

REFERENCES

Barrera-Saldana, H., Takahashi, K., Vigneron, M., Wildeman, A., Davidson, L. and Chambon, P.: All six GC-motifs of the SV40 early upstream element contribute to promoter activity in vivo and in vitro. EMBO J. 4 (1985) 3839–3849.

Basler, K., Oesch, B., Scott, M., Westaway, D., Walchli, M., Groth, D.F., McKinley, M.P., Prusiner, S.B. and Weissmann, C.: Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. Cell 46 (1986) 417–428.

Beckendorf, S.K. and Kafatos, F.C.: Differentiation in the salivary glands of *Drosophila melanogaster*: characterization of the glue proteins and their developmental appearance. Cell 9 (1976) 365–373.

Beutler, E.: Glucose-6-phosphate dehydrogenase deficiency. In Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.C. and Brown, M.S. (Eds.), The Metabolic Basis of Inherited Disease, 5th ed., McGraw-Hill, New York, 1983, pp. 1629–1653.

Cavener, D.R.: Comparison of consensus sequences flanking translational start sites in *Drosophila* and vertebrates. Nucl. Acids Res. 15 (1987) 1353–1361.

Dynan, W.S., Sazer, S., Tijan, R. and Schimke, R.T.: Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature 319 (1986) 246–248.

Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P.: The complete nucleotide sequence of the chick α-actin gene and its evolutionary relationship to the actin gene family. Nucl. Acids Res. 10 (1982) 3861–3876.

Ganguly, R., Ganguly, N. and Manning, J.E.: Isolation and characterization of the glucose-6-phosphate dehydrogenase gene of *Drosophila melanogaster*. Gene 35 (1985) 91–101.

Grima, B., Lamouroux, A., Boni, C., Julien, J., Javoy-Agid, F. and Mallet, J.: A single gene encoding multiple tyrosine hydroxylases with different predicted functional characteristics. Nature 326 (1987) 707–711.

Hori, S.H., Akasaka, M., Ito, H., Hanaoka, T., Tanda, S., Ohtsuka, E., Miura, K., Takahashi, T. and Tang, J.J.N.: Cloning of the glucose-6-phosphate dehydrogenase gene of *Drosophila melanogaster* using 17-base oligonucleotide mixture as probes. Jpn. J. Genet. 60 (1985) 455–463.

Hughes, M.B. and Lucchesi, J.C.: Genetic rescue of a lethal null activity allele of 6-phosphogluconate dehydrogenase in *Drosophila melanogaster*. Science 197 (1977) 1114–1115.

Ingolia, T.D. and Craig, E.A.: Primary sequence of the 5′ flanking regions of the *Drosophila* heat shock genes in chromosome subdivision 67B. Nucl. Acids Res. 9 (1981) 1627–1642.

Karess, R.E. and Rubin, G.M.: Analysis of P transposable element functions in *Drosophila*. Cell 38 (1984) 135–146.

Keller, E.B. and Noon, W.A.: Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs. Nucl. Acids Res. 13 (1985) 4971–4981.

Kozak, M.: Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. 12 (1984) 857–872.

LeFevre Jr., G.: A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In Ashburner, M. and Novitski, E. (Eds.), The Genetics and Biology of *Drosophila*. Academic Press, New York, 1976, pp. 31–66.

Lee, C.-Y., Langley, C.H. and Burkhart, J.: Purification and molecular weight determination of glucose-6-phosphate dehydrogenase and malic enzyme from mouse and *Drosophila*. Anal. Biochem. 86 (1978) 697–705.

Levis, R., Hazelrigg, T. and Rubin, G.M.: Separable *cis*-acting control elements for expression of the *white* gene of *Drosophila*. EMBO J. 4 (1985) 3489–3499.

Lucchesi, J.C. and Rawls, J.M.: Regulation of gene function: a comparison of enzyme activity levels in relation to gene dosage in diploids and triploids of *Drosophila melanogaster*. Biochem. Genet. 9 (1973) 41–51.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'urso, M. and Persico, M.G.: Structural analysis of the X-linked gene encoding human glucose-6-phosphate dehydrogenase. EMBO J. 5 (1986) 1849–1855.

Masters, J.N. and Attardi, G.: Discrete human dihydrofolate reductase gene transcripts present in polysomal RNA map with their 5′ ends several hundred nucleotides upstream of the main mRNA start site. Mol. Cell. Biol. 5 (1985) 493–500.

Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T.: Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA 81 (1984) 2147–2151.

Messing, J.: New M13 vectors for cloning. Methods in Enzymol. 101 (1983) 20–78.

Mitchell, P.J., Carothers, A.M., Han, J.H., Harding, J.D.E., Venolia, L. and Chasin, L.A.: Multiple transcription start sites, DNase 1-hypersensitive sites, and an opposite-strand exon in the 5′ region of the CHO *dhfr* gene. Mol. Cell. Biol. 6 (1986) 425–440.

Mount, S.M.: A catalogue of splice junction sequences. Nucl. Acids Res. 10 (1982) 459–472.

Muskavitch, M.A. and Hogness, D.S.: An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell 29 (1982) 1041–1051.

Nero, D.: The Pentose Phosphate Shunt in *Drosophila melanogaster*: Studies on Dysgenic Mutants of Glucose-6-phosphate Dehydrogenase and 6-Phosphogluconate Dehydrogenase.

Ph. D. Thesis, Cornell University, Ithaca, NY, 1987.

Patel, P.I., Framson, P.E., Caskey, C.T. and Chinault, A.C.: Fine structure of the human hypoxanthine phosphoribosyltransferase gene. Mol. Cell. Biol. 6 (1986) 393–403.

Persico, M.G., Viglietto, G., Martini, G., Toniolo, D., Paonessa, G., Moscatelli, C., Dono, R., Vulliamy, T., Luzzatto, L. and D'urso, M.: Isolation of human glucose-6-phosphate dehydrogenase (G6PD) cDNA clones: primary structure of the protein and unusual 5′ non-coding region. Nucl. Acids Res. 14 (1986) 2511–2522.

Pirrotta, V., Steller, H. and Bozzetti, M.P.: Multiple upstream regulatory elements control the expression of the *Drosophila white* gene. EMBO J. 4 (1985) 3501–3508.

Pustell, J. and Kafatos, F.C.: A high speed, high capacity homology matrix: zooming through SV40 and polyoma. Nucl. Acids Res. 10 (1982) 4765–4782.

Pustell, J. and Kafatos, F.C.: A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. Nucl. Acids Res. 12 (1984) 643–655.

Reynolds, G.A., Goldstein, J.L. and Brown, M.S.: Multiple mRNAs for 3-hydroxy-3-methylglutaryl coenzyme A reductase determined by multiple transcription initiation sites and intron splicing sites in the 5′-untranslated region. J. Biol. Chem. 260 (1985) 10369–10377.

Roebuck, K.A. and Stumph, W.E.: The 5′-flanking DNA of chicken U1 RNA genes shares certain characteristics of 'housekeeping' gene promoter regions. DNA 4 (1985) 86.

Rubin, G.M. and Spradling, A.C.: Genetic transformation of *Drosophila* with transposable element vectors. Science 218 (1982) 348–353.

Rubin, G.M. and Spradling, A.C.: Vectors for P element-mediated gene transfer in *Drosophila*. Nucl. Acids Res. 11 (1983) 6341–6351.

Takizawa, T., Huang, I., Ikuta, T. and Yoshida, A.: Human glucose-6-phosphate dehydrogenase: primary structure and cDNA cloning. Proc. Natl. Acad. Sci. USA 83 (1096) 4157–4161.

Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, T.M., Van der Voorn, L., Van Ormondt, H. and Van der Eb, J.A.: Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. 4 (1985) 437–443.

Williamson, J.H. and Bentley, M.M.: Comparative properties of three forms of glucose-6-phosphate dehydrogenase in *Drosophila melanogaster*. Biochem. Genet. 21 (1983) 1153–1166.

Yoshida, A. and Beutler, E. (Eds.): Glucose-6-Phosphate Dehydrogenase. Academic Press, New York, 1986.

Communicated by A.D. Riggs.